Emerson and the Digital Humanities

Amy Earhart

When teaching Emerson we are faced with the challenge of giving students skills with which to engage with Emerson's seemingly dense and difficult writing. Through careful use of digital humanities tools, we might guide students through exercises that provide access points to Emerson's work, aid in interpretation and evaluation of Emerson's ideas, and spur classroom discussion. To illuminate how digital pedagogy might enhance the study of Emerson in the classroom, what follows are sample exercises for classroom use that will help students 1) develop an understanding of textual accuracy and transmission, and 2) use digital methods to analyze Emerson's work. Though focused on specific tools and techniques, these examples highlight the way by which digital pedagogies deepen student engagement with and understanding of literary texts.

One of the concerns of scholars is that students do not understand the importance of textual accuracy and transmission. To address this concern, I have used the *Juxta Commons* digital collation tool to help students understand why reliable texts are crucial to scholarly inquiry. Jerome McGann's brainchild, *Juxta* was developed under the auspices of Nineteenth-Century Scholarship Online (NINES) and allows the user to compare versions of texts (collation, a classic technique in editing) and to complete various tasks related to digital edition building. A free, Web-based tool, *Juxta* is designed to create "interpretive comparisons" that the user might then examine to fulfill "the basic humanities interpretive act: critical comparison" (McGann). The use of *Juxta* in the classroom teaches students to see the impact of textual reliability on the way that we understand and interpret Emerson. More simply put, a flawed literary text will produce a problematic, even inaccurate scholarly analysis. A comparison of various editions or versions of an Emerson essay allows the instructor to emphasize the importance of textual variability and to create points of engagement for student analysis of the text, what Amanda Gailey calls "attentive reading."

The first challenge instructors face is to identify reliable electronic versions of Emerson's writings. Emerson's works have been digitized at a rapid pace, though the quality of texts is varied. Much of our current digital textual data is limited because of such transcription issues. Large-scale projects such as Google Books or Eighteenth Century Collections Online (ECCO) use optical character recognition (OCR), the translation of images to machine-readable files, of variable quality. This impacts not only the readability of materials but also the searchability, as most full-text searches of large literary data sets are based on OCR'd documents—an issue well documented by critics including Geoffrey Nunberg, who has been outspoken in his critique of misnaming, misdating, and other metadata errors in *Google Books*. Since Harvard's *The Collected Works of Ralph Waldo Emerson*, the most trusted edition of Emerson's work, is not available in manipulable digital format, I have turned to the Ralph Waldo Emerson Society's Web site, which includes links to higher-quality digital transcriptions of Emerson texts on its "Related Sites" page. When teaching Emerson's "New England Reformers," for example, the instructor could have students compare two different versions of the essay: an 1844 essay, such as the version from Jone Johnson Lewis's *Emerson Central*, to a Centenary Edition essay, such as in the University of Michigan's digitized *The Complete Works of Ralph Waldo Emerson*. Given critical concerns regarding Edward Waldo Emerson's editing of his father's texts, the comparison of the two versions makes for interesting class discussion and reflection regarding the way that editing choices impact the texts we study.

Once the primary texts are located, using *Juxta* is relatively simple. Students create a free account through the Juxta Commons Web site and then upload a digital file of the Emerson text or texts. *Juxta* accepts a wide variety of file formats (html, doc, and pdf), making it a versatile program as the instructor doesn't have to spend a great deal of time producing documents for analysis. When the digital versions are uploaded into *Juxta* they must be converted into "witnesses" and grouped into comparison sets, simple tasks to complete with the easy-to-use tool. Once the texts are converted and grouped, students may begin to use the visualizations created by *Juxta* to compare the 1844 and Centenary versions of Emerson's "New England Reformers." *Juxta* creates a heat map (textual differences suggested by gradations of color), a side-by-side visualization, a two-text comparison with lines linking differences in texts, and a histogram (a visualization of the overall rate of change). The first two visualizations prove most helpful for pedagogical purposes. The heat map of the two witnesses of "New England Reformers" displays several textual differences (see "Heat Map," dx.doi.org/10.17613/M68D7D).[1]

In this particular paragraph the differences expressed are primarily punctuation and capitalization. For a closer look at the differences, students may use the side-by-side comparison (see "Side-by-Side Comparison," dx.doi.org/10.17613/M64M43). As they look through the text, students will notice that the witnesses include word choice differences. For example, Emerson writes that the university is either "ludicrously styled" or "ludicrously called." Such substantive differences will surprise and engage students and allow the instructor to emphasize that even small editorial changes might impact the way we understand a text. Such "interactivity and visuals" in digital pedagogy approaches, argues W. Michele Simmons, "increase opportunities for learning but also simultaneously introduce complexities."

Students will also be interested in the discrepancies in document structures, including the paragraph breaks, spacing, and punctuation. For some scholars, such as Peter Shillingsburg, the issue of digital accuracy is central but also revealing: "Error can be just noise. We correct it or ignore it just as we talk and listen over the noise of a passing train or when surrounded by cocktail chatter. Error can also be serendipitous discovery, leading us to new insights and

allowing for repurposing of old texts" (161). Student manipulation of texts, then, helps them to engage in the important conversations about literature, texts, and the creation of information in the digital age.

While these exercises hint at the importance of a reliable text, assignments explicitly designed to require students to examine textual editing are also productive. Some critics, such as Gailey, situate such concerns in the digital humanities, noting "the continuation into the digital age of the disciplines concerned with the materiality and representation of texts, such as textual studies and editorial theory" (192). A *Juxta* assignment that allows students to locate corruptions in an Emerson text not only emphasizes the materials and representations of texts, but it reminds students of the importance of textual accuracy. With a digital collation project students learn "to be attentive to small details, discovering that changes in small details—even when not authorized by the author—can alter dramatically interpretation" (Hawkins 138). For this assignment, I choose two different versions of the same text: Lewis's carefully edited 1844 text of "New England Reformers" and the same essay from the 1909 Harvard Classics series, found on the *Internet Archive* that has been machine transcribed from the *Google Books* digitization (English Traits). I ask students to think about what they might find when they collate the two versions. After the first exercise, students recognize that the two versions will have differences, but they are unaware of the differences introduced by the machine transcription. When collated, students learn that there are numerous substantive differences beyond the two differing versions of the essay (see "Collation Heat Map," dx.doi.org/10.17613/M60X1T).

In the exercise students realize that the file title is incorrectly translated ("Ralph W,lUX) Emer5on") and individual words are incorrectly rendered. Such inaccuracies cause great difficulties with text mining and with full-text searches. The distortion of *hospital*, for example, would mean that a search looking at the word choices Emerson uses in his entire corpus that reference disease would miss this important reference. Important when teaching with digital techniques is to reveal the limitations of the analysis that might be made with technological innovation. Side-by-side comparisons reveal that the primary differences in the text are errors introduced by the machine transcription (see "Collation Side-by-Side Comparison," hcommons.org/deposits/item/hc:13665).

While collation tools are useful to engaging students in understandings of textual accuracy and transmission, the use of topic-modeling software provides additional teaching possibilities. Topic modeling is described by Miriam Posner as "a method for finding and tracing clusters of words (called 'topics' in shorthand) in large bodies of texts." Topic modeling is central to digital humanities work, and student familiarity with such an approach to textual analysis is increasingly a centerpiece of digital pedagogy. Though there are numerous tools that might be of use, Stefan Sinclair and Geoffrey Rockwell's Web-based plug-and-play set of tools, *Voyant*, has long been my go-to for undergraduate students. *Voyant* is fairly easy to use and produces high-quality visualizations for analysis. In addition, the tools are currently being revised, a fairly unique occurrence among start-up digital humanities tools, which often run out of funding and are often unmaintained, unstable, and unacceptable for classroom applications. *Voyant*'s new updates should reassure instructors that assignments developed for the *Voyant* tools will be possible in the future.

Topic modeling allows students to visualize large-scale patterns across multiple texts. Paul Fyfe describes this as a sort of "'hybrid' critical work" that "allows for specificities within close contexts as students read, and for connections within and across texts." In a course where students are reading multiple Emerson essays, for example, topic modeling might aid students in understanding his shifting ideas over time. In a class focused on Emerson and slavery, for instance, the instructor might have students upload three related Emerson texts to *Voyant* for comparison: "Address on Emancipation in the British West Indies" (1844), "The Fugitive Slave Law" (1854), and "The Emancipation Proclamation" (1862) (see "Cirrus Word Cloud," dx.doi.org/10.17613/M6RD6C).[2]

Once the texts are uploaded, *Voyant* performs multiple visualizations using a suite of tools. The Cirrus tool produces a word cloud that represents the number of times a word appears in the text, which allows the instructor to ask the students to interpret Emerson's repetition of certain words. Because Cirrus highlights Emerson's use of *man* and *men* as the most used words in these three Emerson essays, the instructor might prompt students to place Emerson's essays in the context of other literature of the nineteenth century, such as contemporaneous slave narratives, including *Narrative of the Life of Frederick Douglass*, that reclaim *man* for African Americans. Discussion of the repetitive words used by Emerson, such as *man*, *power*, *good*, and *slavery*, might also prove illuminating, revealing to students "that they are creating knowledge and creating it through acts of interpretation" (Hawkins 139).

Voyant's Corpus Text tool provides additional information for such discussions. When multiple documents are added to *Voyant*, the Corpus Text tool shows the trend of individual words between documents (see "Corpus Terms," dx.doi.org/10.17613/M6MM5T). Since the three essays have been uploaded in progressive order, trends across time and essay are revealed. When students examine the visualizations they might posit that Emerson's use of *slave* might have dropped between 1844 and 1862 because of the Emancipation Proclamation (1863). Or that the use of *man*, which peaks in usage in Emerson's "The Fugitive Slave Law," reveals how Emerson understands slavery in relationship to the individual. The tool's ability to reveal such patterns, then, allows the instructor to encourage the humanistic interpretations that are the basis of literary analysis.

In addition to tracking individual word repetition throughout multiple texts, *Voyant*'s Links tool encourages students to think about what concepts of the texts are interrelated. Links reveals that Emerson uses the term *men* in relationship to a host of words, including *colleges*, *England*, *accomplished*, *law*, *class*, etc. (see "Links," dx.doi.org/10.17613/M6GX0S). What is Emerson saying about men in college? Why the connection to England or to law? Students

might begin by hypothesizing what Emerson might be suggesting by the relationship and, after initial discussion, return to the essay. If students thought that Emerson was interested in sorting out men according to class status, then we might locate the passage from "The Fugitive Slave Law": "For every man speaks mainly to a class whom he works with and more or less fully represents. It is to these I am beforehand related and engaged, in this audience or out of it—to them and not to others" (*Complete Works* 11: 218 [U of Michigan]). Reflecting on the association of *class* and *man* might lead a student to revise his or her opinion of Emerson's ideas about class. While the movement between tool (a distant reading of the text) and text (a close reading) may not provide a brilliant aha moment, using visualizations does ask the student to engage in the text in a deeper and more interactive method than would be easily accomplished in a traditional lecture class. Students feel far more invested in their interpretation of the text in such an approach, leading to a more engaged discussion of Emerson.

The use of digital pedagogical techniques within the classroom allows the instructor to provide a window into what students perceive as the dense and difficult language of Emerson. Through the use of carefully designed assignments, students might find Emerson's texts challenging puzzles for interpretation. The digital tools that I have discussed are open access, easy to use, and applicable to varying pedagogical approaches and classrooms. The tools allow us to emphasize the best practices in our field: interpretation through distant and close reading, contextual interpretation, and attention to the transmission of a document.

NOTES

[1] All figures are stored at the MLA commons CORE repository (mla.hcommons.org/core).

[2] For this assignment, I use texts from Lewis's "Ralph Waldo Emerson Texts" Web page.